CreditOne Customer Default Identification Report

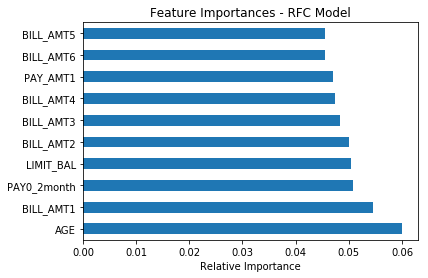
CreditOne wants to resolve an issue regarding an increase in the default rate of its customers. They have addressed this issue previously with the following findings:

* We cannot control customer spending habits.
* Finding do not always point to the underlying "why".

My initial hypothesis for this study was the most likely candidates to default would be single women, aged mid to late 20’s, with a university degree who used the card as a revolving charge with a charge limit of 50K. This initial hypothesis was based solely upon the characteristics of the samples.

Models of the data were designed and tested to determine the most import aspect for CreditOne in who the most likely candidates are for future defaults.

The first pass at the analysis, I tried to determine the amount for PAY\_AMT6 as an indicator of those who would default in the future. The results pointed me in the direction of the attribute BILL\_AMT6. This field represents the amount of the customer’s bill on the six month of data capture. This field, though significant in determining default status, was not convincingly significant (14% accuracy). The best models were wildly inaccurate.

Next, I re-tuned the analysis and tried a different approach; using different modeling techniques to predict who would be most likely to default in the future. I added demographic information not used for the first-pass analysis to the model. This time around the model identified AGE, BILL\_AMT1, PAY0\_2month (those who were two months behind in payments), and LIMIT\_BAL as the most indicative (99.97% accuracy). This result would support the idea the most likely people to default are those with the least amount of credit experience. Consumer Education can be a possible remedy. The results of the model are stunning, but may be usually high. Further analysis is warranted before making a decision.

TRAINING RESULTS

Cross Validation Score: [0.82123304 0.81761905 0.81071429 0.80809524 0.81829007]

Mean Accuracy: 0.99971

TESTING RESULTS

Mean squared error: 0.18: (Lower numbers are better)

R Squared of testing: 1.00: (Higher numbers are better, but be careful of overfitting)

Classification report:

precision recall f1-score support

0 0.84 0.95 0.89 7060

1 0.65 0.37 0.47 1940

accuracy 0.82 9000

macro avg 0.75 0.66 0.68 9000

weighted avg 0.80 0.82 0.80 9000

Kappa: 0.3710869

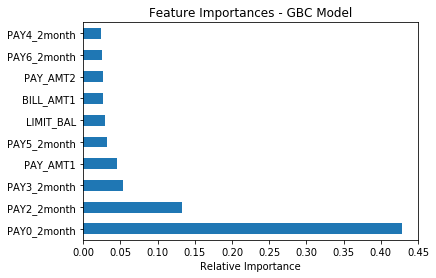
Confusion matrix (Future DEFAULT):

No Yes

[[6682 378]

[1231 709]]

Another model was deployed to verify the newly defined assumption. Using the same data but a different algorithm, similar results should be expected. The results should still be in agreement with the previous assumption.

This model removed AGE as a factor and added those who are chronically late in paying their bill. Attributes PAY0\_2month, PAY2\_2month, PAY3\_2month, and PAY\_AMT1 were the most significant (82.45% accuracy). This embraces the idea of Consumer Education and further gives value to the idea of a probationary period for new accounts; 6 months before approving an increase to the LIMIT\_BAL.

TRAINING RESULTS

Cross Validation Score: [0.81285714 0.82666667 0.80928571 0.81738095 0.8202381 ]

Mean Accuracy: 0.82448

TESTING RESULTS

Mean squared error: 0.18: (Lower numbers are better)

R Squared of testing: 0.82: (Higher numbers are better, but be careful of overfitting)

Classification report:

precision recall f1-score support

0 0.84 0.95 0.89 7060

1 0.66 0.36 0.47 1940

accuracy 0.82 9000

macro avg 0.75 0.65 0.68 9000

weighted avg 0.80 0.82 0.80 9000

Kappa: 0.3707693

Confusion matrix (Future DEFAULT):

No Yes

[[6704 356]

[1242 698]]

A third, more processing intensive model, was used to identify only those attributes which mathematically contribute to the attribute DEFAULT. It identified in no particular order the following attributes. The results are not surprising but still demonstrates AGE, LIMIT\_BAL, and the amounts charged and ability to pay are contributing factors. Consumer Education is still a viable remedy.

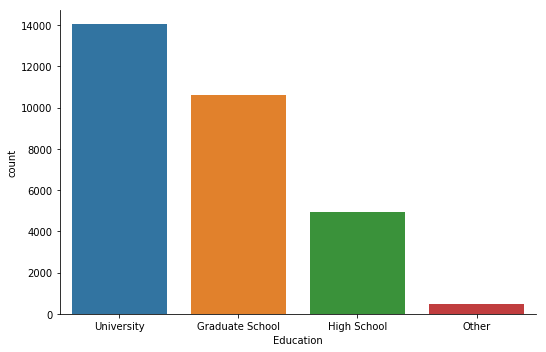
Number of Features: 15

|  |  |
| --- | --- |
| Attribute | Ranking |
| LIMIT\_BAL | 1 |
| AGE | 1 |
| BILL\_AMT1 | 1 |
| BILL\_AMT2 | 1 |
| BILL\_AMT3 | 1 |
| BILL\_AMT4 | 1 |
| BILL\_AMT5 | 1 |
| BILL\_AMT6 | 1 |
| PAY\_AMT1 | 1 |
| PAY\_AMT2 | 1 |
| PAY\_AMT3 | 1 |
| PAY\_AMT4 | 1 |
| PAY\_AMT5 | 1 |
| PAY\_AMT6 | 1 |
| PAY0\_revolving | 1 |

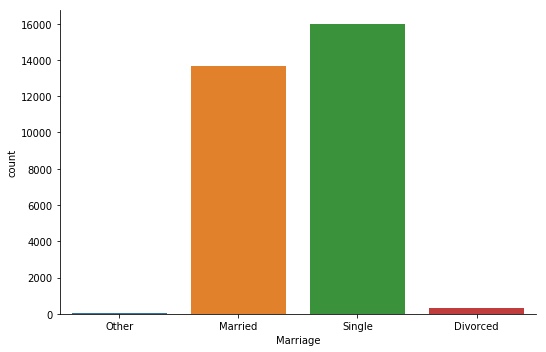
Mean Accuracy: 0.77633

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Training Model Scores | | Prediction Scores | | | | |  |
|  | Accuracy | RMSE | R2 | Accuracy | Support | Kappa | |
| RFC | 0.99971 | 0.18 | 1.00 | 0.82 | 9000 | 0.3684304 | |
| GBC | **0.82448** | **0.18** | **0.82** | **0.82** | **9000** | **0.3707693** | |
| SVC | 0.77633 | 0.22 | 0.78 | 0.78 | 9000 | 0 | |
| PCA (SVC) | 0.80389 | 0.20 | 0.80 | 0.80 | 9000 | 0.2518876 | |

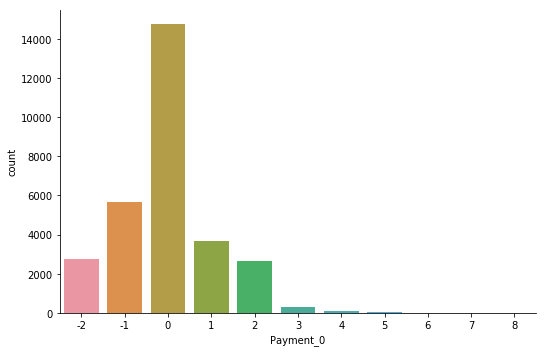
In summary, the data provided does not point to any particular attribute as being the most significant factor in determining who is likely to default in the future. But it does demonstrate multiple factors combined can contribute to defaulting on an account. A remedy to this issue will need to encompass the attributes identified. If a course of action is to provide credit counseling courses with a 6 month freeze on Credit Limit increases, customers may be initially resistant. Offering worthwhile incentives can help alleviate any misgivings. In the long view, CreditOne wants to avoid churn of customers and not lose money due to defaulted accounts.



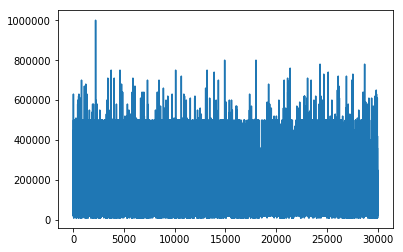
Education



Marital Status



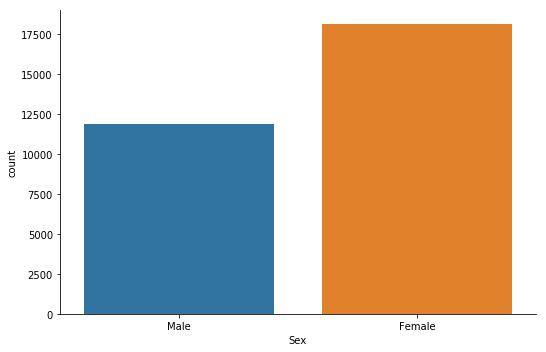
Account Type



Credit Limit



Age



Gender

